

Sentiment analysis algorithms  
for the Belarusian NooJ module  
in touristic sphere

# Co-authors



Yuraś Hiecevič



Anastasija Drahun



Aliena Kryvałcevič



Jaŭhienija Zianoŭka



Nastassia Kazloŭskaja



Aliaksandr Ščarbakoŭ

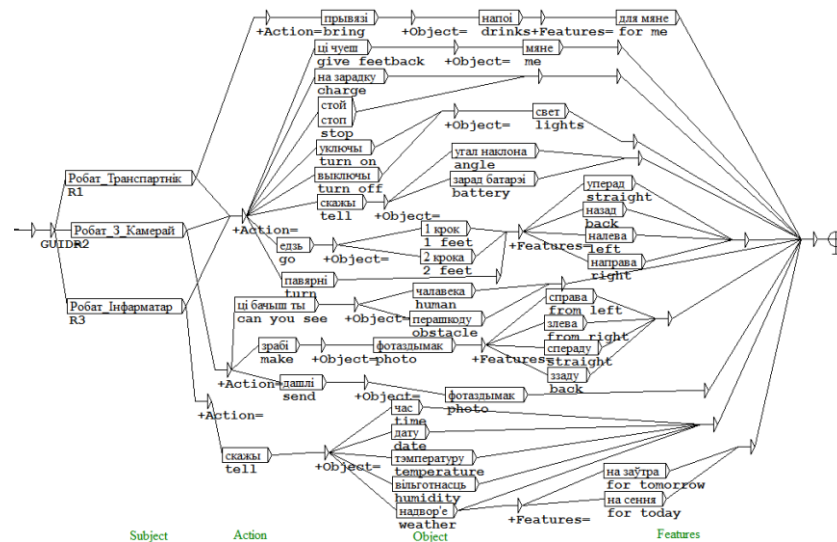
# Background of Belarusian NooJ module 2011-2015

- 2011 Overview of Belarusian and Russian grammatical dictionaries
- 2012 First Belarusian, Russian \*.nod dictionaries, publication of Belarusian NooJ module with texts, projects, grammars, dictionaries
- 2013 Syntax grammars for QEMU and direct speech processing
- 2014 Next steps in syntax grammars for QEMU processing
- 2015 First morphological and syntax grammars for
  - **syllabification,**
  - **phonemic**
  - **and prosodic transcriptions.**
  - Foundation of Belarusian NooJ team
  - Minsk NooJ conference

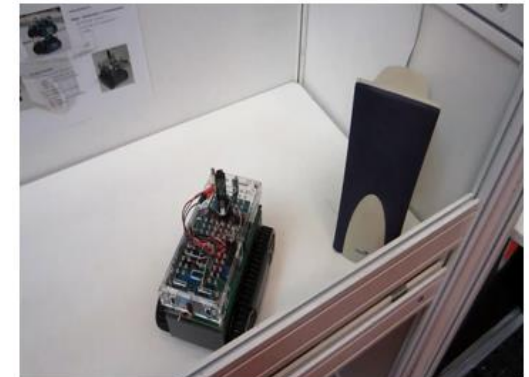


# Background of Belarusian NooJ module 2016 (first half)

- 2016 Next steps in morphological and syntax grammars for syllabification, phonemic and prosodic transcription;
  - generating of texts for robotics command support;



Robot & Belarusian TTS



# Background of Belarusian NooJ module 2016 (second half) – 2017

- **Weekly news letters** for Belarusian NooJ team and Max Silberstein about done tasks and tasks that we are going to do
  - Hanna Stanislavenka
- Belarusian video **tutorials for NooJ**
  - Julia Baradzina, Maryna Marchyk
- **Sentiment analysis** for Belarusian touristic sphere
- **New Java NooJ core** implementation
  - Alaksandar Scharbakou, Yuras Hetsevich





# Belarusian video tutorials for NooJ

# 12 videos in Belarusian for new NooJ participants

The screenshot displays a YouTube playlist titled "NooJ tutorials: па-беларуску" (NooJ tutorials: in Belarusian) from the channel "Speech synthesis and recognition laboratory". The playlist contains 12 videos, with the first one, "Усталёўка праграмы" (Program installation), selected. The video player shows a woman speaking. The interface is in Belarusian. The Windows taskbar at the bottom shows the Start button, taskbar search, and several open applications including "SA\_NOOJ\_A... .ppbx", "пунк", "(32) Nooj tut...", "Total Comma...", and "Microsoft Pow...".

- [https://www.youtube.com/playlist?list=PLtc\\_R9i0zr6QjyLk5\\_Vn\\_9F4balHK42XW](https://www.youtube.com/playlist?list=PLtc_R9i0zr6QjyLk5_Vn_9F4balHK42XW)



Sentiment analysis algorithms  
for the Belarusian NooJ module  
in touristic sphere



# Plan in general

- Overview of the problem and task definition
- Corpora collecting: training and test
- Experiments: 1<sup>st</sup>, 2<sup>nd</sup> on basis of training corpora
- Evaluations: on basis of training corpora and final testing corpus
- Future plans



# Overview of the sentiment analysis (SA) problem, focusing to Belarus

- Sentiment analysis aims to determine the opinion of a speaker, writer, or other subject with respect to some topic (*emoticons “)))”*, *good, bad, sad words about object*)
- Belarus becomes more welcome and open for touristic flow
- **How can we measure opinion of newcomers about Belarus?**



There is no any SA software  
for Belarusian touristic content...

(:



We hope that  
NooJ can help us!  
:)



# Task definition

- Sphere – tourism
- Main topic of reviews – Mirski castle (Мірскі замак)
- Sources – Internet pages and social networks hash tags with reviews for main topic





# Examples of reviews

## POSITIVE

- *«Замок очень красивый и внутри, и снаружи. Впечатляющие виды днем и вечером с подсветкой. Рядом озеро и небольшой парк с усыпальницей»*

## NEGATIVE

- *«При въезде в город сам замок не виден, указателей никаких нет. Немножко заплутали, но благо город небольшой, поэтому быстро нашли. Конечно замок живую не производит величественного впечатления. Но при покупке билета нам не было предложено аудиогuida. (Обязательно просите, потому что без него или без экскурсовода там нечего делать). Экспозиция очень слабенькая.»*



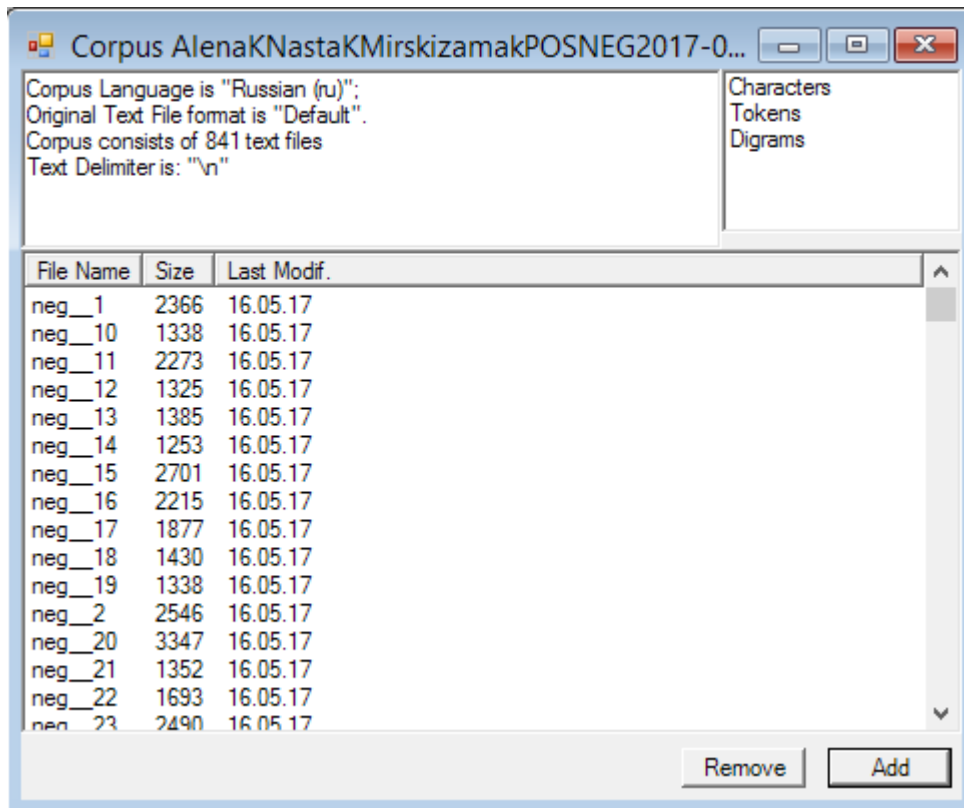
# Corpora collecting (txt files)

- 1st training corpus
  - pos. reviews – 503
  - neg. reviews – 168
- 2nd training corpus
  - Additional reviews were added to the 1<sup>st</sup> training corpus
  - pos. reviews – 634
  - neg. reviews – 207
- Test corpus
  - Reviews are not appear in training corpora
  - pos. reviews – 83
  - neg. reviews – 24





# Transferring training and test corpora into NooJ corpora



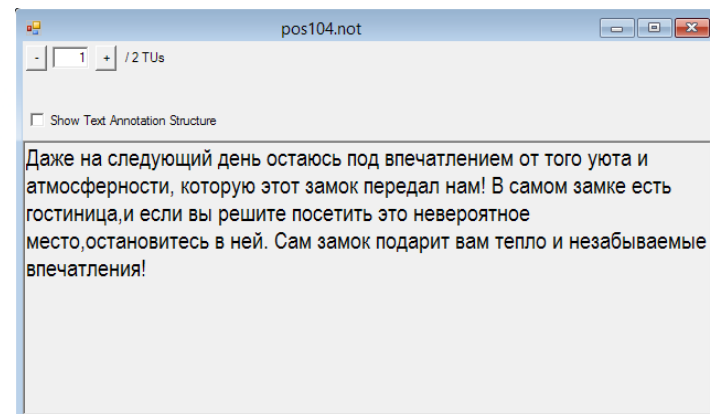
Corpus AlenaKNastaKMirskizamakPOSNEG2017-0...

Corpus Language is "Russian (ru)";  
Original Text File format is "Default".  
Corpus consists of 841 text files  
Text Delimiter is: "\n"

Characters  
Tokens  
Digrams

File Name	Size	Last Modif.
neg__1	2366	16.05.17
neg__10	1338	16.05.17
neg__11	2273	16.05.17
neg__12	1325	16.05.17
neg__13	1385	16.05.17
neg__14	1253	16.05.17
neg__15	2701	16.05.17
neg__16	2215	16.05.17
neg__17	1877	16.05.17
neg__18	1430	16.05.17
neg__19	1338	16.05.17
neg__2	2546	16.05.17
neg__20	3347	16.05.17
neg__21	1352	16.05.17
neg__22	1693	16.05.17
neg__23	2490	16.05.17

Remove Add

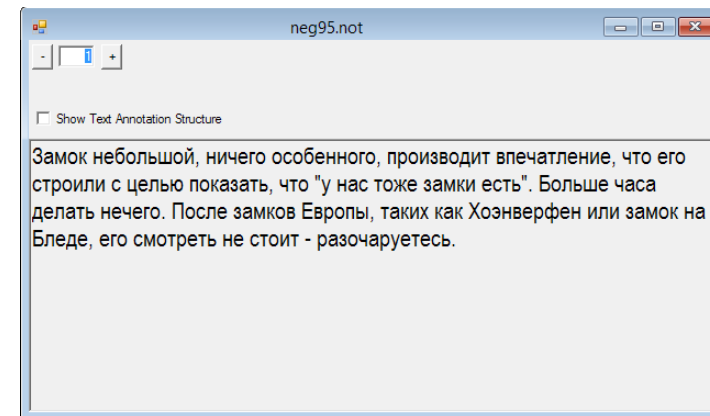


pos104.not

1 / 2 TUs

Show Text Annotation Structure

Даже на следующий день остаюсь под впечатлением от того уюта и атмосферности, которую этот замок передал нам! В самом замке есть гостиница, и если вы решите посетить это невероятное место, остановитесь в ней. Сам замок подарит вам тепло и незабываемые впечатления!



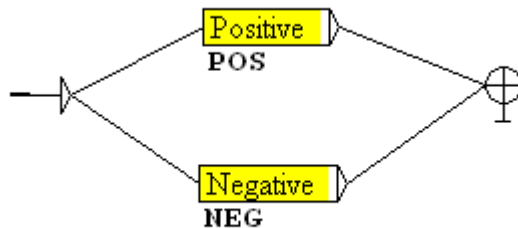
neg95.not

Show Text Annotation Structure

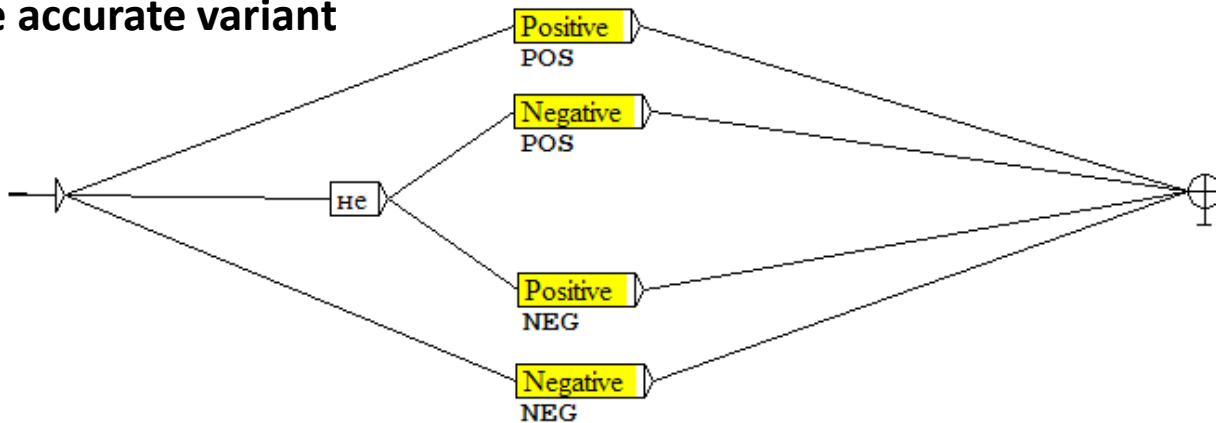
Замок небольшой, ничего особенного, производит впечатление, что его строили с целью показать, что "у нас тоже замки есть". Больше часа делать нечего. После замков Европы, таких как Хоэнверфен или замок на Блеме, его смотреть не стоит - разочаруетесь.

# Experimental grammars under the basis of the 1<sup>st</sup> training corpus

Naïve variant



More accurate variant



# One of the “Positive” subgraph

Positive phrases (true answers) – 348

The screenshot shows a software interface with two windows. The left window, titled "Structure of Pos\_Neg\_...", displays a tree view of a syntactic grammar. The right window, titled "Pos\_Neg\_all\_in\_one\_2017-05-15\_1148\_Alena.nog", displays a list of positive phrases.

**Structure of Pos\_Neg\_...**

- Expand
- Collapse
- Refresh
- Main
  - AdditionalNegative
  - AdditionalPositive
  - Negative
  - Positive
    - ... ) (X)
    - ... ) (X)

**Pos\_Neg\_all\_in\_one\_2017-05-15\_1148\_Alena.nog**

Russian /Russian syntactic grammar consists of 5 graphs.

Positive

Архитектура завораживает  
антураж  
аутентичность  
аутентичные  
Более приятное впечатление  
Большая коллекция  
Буду рекомендовать  
Билеты недорогие  
безумно понравилось  
богатой историей  
большая территория  
браво  
Великолепен вид  
Великолепный  
Восстановлен хорошо  
Восхитительное место  
Впечатляющие виды  
Все дышит историей  
Все очень мило  
веет стариной  
великолепен  
великолепная  
великолепное  
великолепном состоянии  
великолепный  
великолепны  
весьма приличный

# One of the “Negative” subgraph

Negative phrases (true answers) - 98

Structure of Pos\_Neg... Pos\_Neg\_all\_in\_one\_2017-05-15\_1148\_Alena.nog

Expand Collapse Refresh

Main

- AdditionalNegative
- AdditionalPositive
- Negative**
- Positive
  - ) (X)
  - ) (X)

Russian /Russian syntactic grammar consists of 5 graphs.

Negative

плохое  
агрессивный персонал  
Безопасность на нуле  
без особого интереса  
бес конца повторялся  
больше лоска, вылезаности  
большие группы туристов  
большие очереди  
бубнения гида  
будет тяжеловато  
бутафорская  
было закрыто  
проблема  
Витые лестницы  
некомфортно  
взбираться на верх сложно  
впечатлила не особо  
врят ли впечатлит  
все на современный лад  
все неинтересно  
выбор небогат  
высокими ступенями  
высосано из пальца  
высосаны из пальца  
в шоке  
Глобальный минус  
гораздо больше туристов  
громоздким и массивным  
гулять негде  
густо восстановленный  
Далеко не все в курсе

# Concordance for the 1st experiments

NooJ - [Concordance for Corpora]

File Edit Lab Project Windows Info CORPUS CONCORDANCE

Reset Display: 5 characters before, and 5 after. Display:  Matches  Outputs

Text	Before	Seq.	After
neg137.not	план. Музей здесь вообще внимания	не достоин/NEG	.
neg138.not	походить внутри. Но это явно	не стоит/NEG	тех денег.
neg139.not		Отреставрированный/POS	- как это часто водит
neg139.not	часто водится в наших краях -	не без вопросов/NEG	.
neg14.not	После Несвижского замка этот оказался	громоздким и массивным/NEG	.
neg140.not	этот тоже не плох поэтому	рекомендую/POS	посещать эти замки
neg141.not	Насколько	красивым/POS	он выглядит снаружи
neg141.not	он выглядит снаружи, настолько же	не интересный/POS	он внутри. Экспозиц
neg141.not	он выглядит снаружи, настолько же	не интересный/NEG	он внутри. Экспозиц
neg141.not	Башни открытые для туристов почти	пустые/NEG	, смотреть не на что.
neg141.not	открытые для туристов почти пустые,	смотреть не на что/NEG	. В музейной части е
neg141.not	и платьям. По ценам немного	дороже/NEG	чем в городе, но не
neg142.not	Экспозиции пока еще	не радуют богатством/NEG	, но заметно, что над
neg142.not	Если не обращать внимания на	явный новодел/NEG	
neg143.not	Производит	не такое яркое впечатление/NEG	как Несвиж, но тем
neg143.not	Несвиж, но тем не менее	посетить обязательно/POS	стоит. Из минусов -
neg143.not	менее посетить обязательно стоит. Из	минусов/NEG	- очень крутые дерев
neg143.not	посетить обязательно стоит. Из минусов -	очень крутые/NEG	деревянные и камен
neg143.not	Подняться на галерею пожилому человеку	трудно/NEG	.
neg144.not	для взрослых - 70 000 бел руб. Отреставрирован	не понятно/POS	- без особой привязк
neg144.not	для взрослых - 70 000 бел руб. Отреставрирован	не понятно/NEG	- без особой привязк
neg145.not	Подниматься на башни	тяжело/NEG	даже для здорового
neg145.not	и молодого человека так как	лестница очень узкая/NEG	Мне показалось цен
neg145.not	лестница очень узкая Мне показалось	цена за билет завышен/NEG	. Народу очень много
neg145.not	показалось цена за билет завышен .	Народу очень много/NEG	.
neg145.not	Экспозиция музея	впечатлила не особо/NEG	, представлена, в о
neg146.not	что замок и его помещения	тщательно/POS	отреставрированы, 1

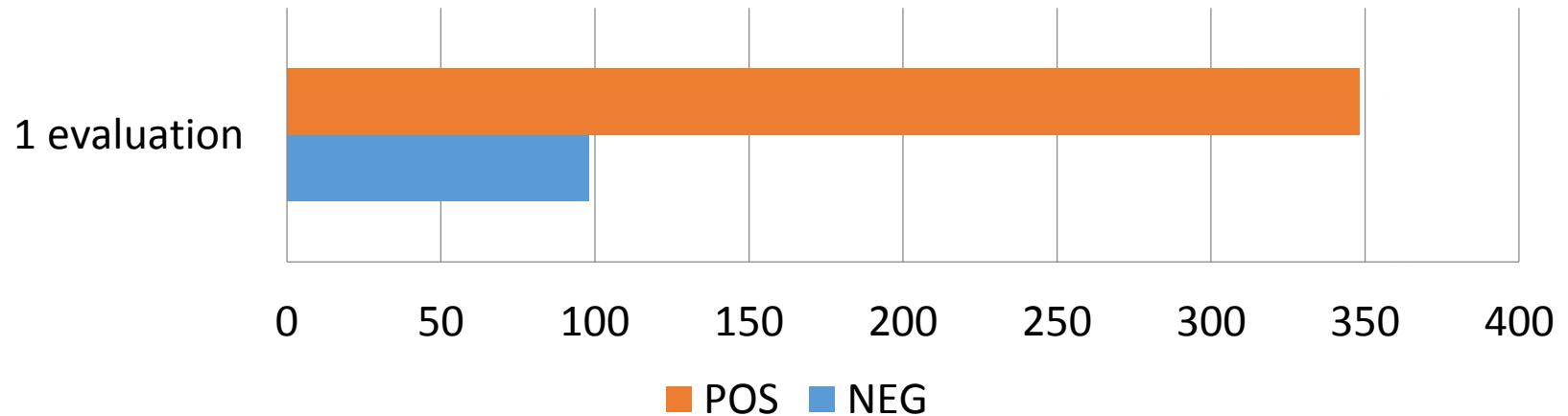
# Evaluation of the 1<sup>st</sup> experiments

- (automatically calculated in the MS Excel)
- Precision: 0,91. Recall: 0,16. F-measure: **0,27.**
- **Discussion – It works! But it is not good accuracy.**
- **Decision – we have to**
  - **update “true answers” because recall is too low**
  - **and check it under 2<sup>nd</sup> corpus for training**

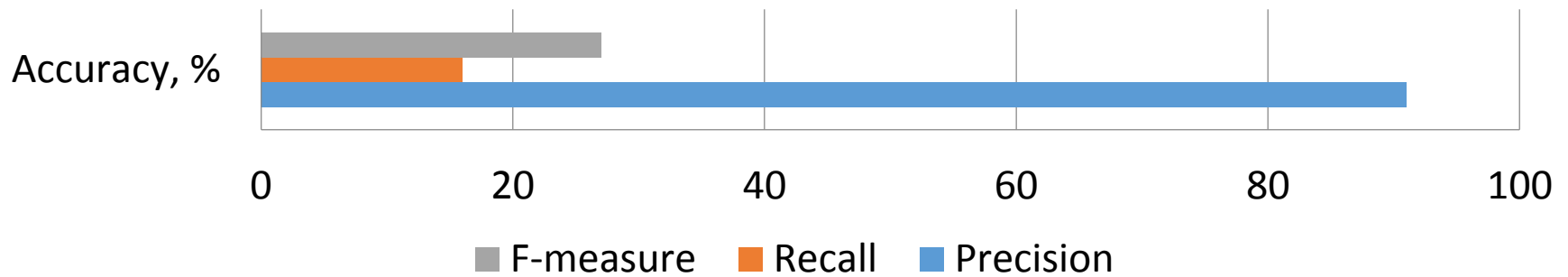
						ВЫДАЛЯЕМ ПАУТОРЫ						
										120	92	18
ДАДЗЕНЬЯ З КАНКАРДАНСА НУДЖ						клеім назву для ўнікальнасьці	выводзім толькі 1 раз усю назву	маркер	хвосьцік	коль сапраўды POS	сапраўды NEG	
neg	11.not	невозможно. Это	минус	NEG	. Причём значитель	neg11.not	neg11.not	neg	15.not	1	0	1
neg	15.not	Внутри располо	очень крутые	NEG	лестницы с	neg15.not	neg15.not	neg	15.not	1	0	1
neg	15.not	краях, то	обязательно пос	POS	Мирский замок	neg15.not				0	0	0
neg	16.not	аудиогид, что	очень удобно	POS	в залах	neg16.not	neg16.not	neg	16.not	1	0	0
neg	16.not	оставить залог.	Проблема	NEG	в том	neg16.not				0	0	0
neg	17.not	одного человека	ступени высокие	NEG	, хорошо, что	neg17.not	neg17.not	neg	17.not	1	0	1
neg	17.not	ступени высокие	хорошо	POS	, что есть	neg17.not				0	0	0

# Charts for the 1<sup>st</sup> experiment data analysis

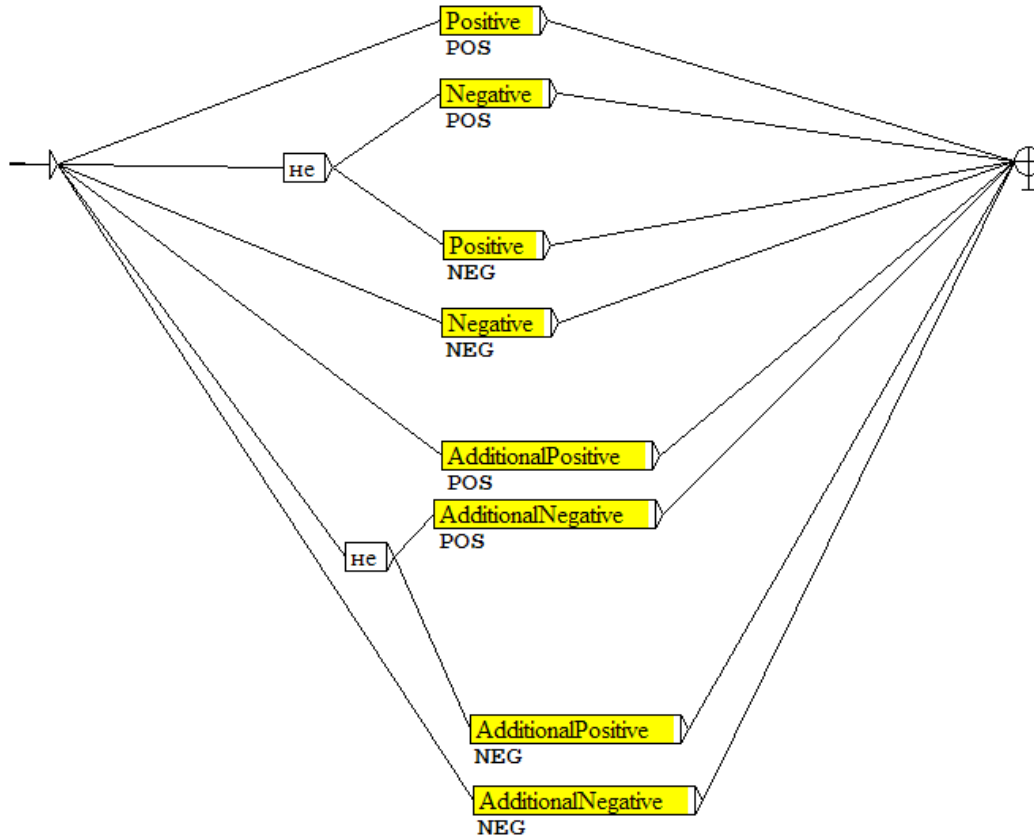
## Amount of "true answers"



## Accuracy



# Experimental grammars on the basis of the 2<sup>nd</sup> training corpus





# True answers for subgraphs of the 2nd training corpus

- Positive phrases – 524
- Negative phrases – 404



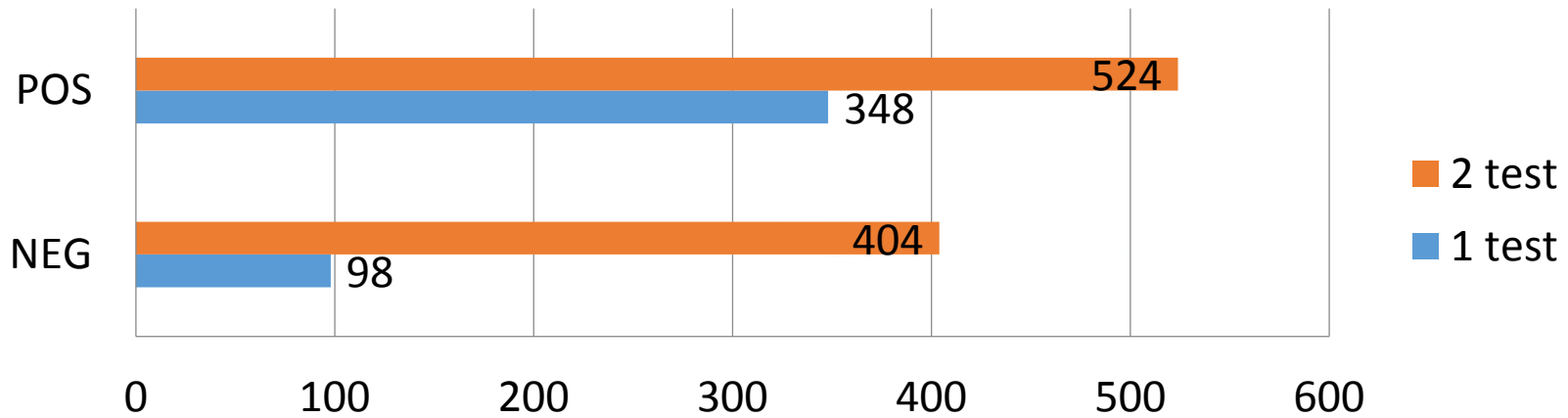
# Evaluation of the 2<sup>nd</sup> experiment

- Precision: 0,92. Recall: 0,89. F-measure: **0,90**
- **Discussion – It is good accuracy! Can we use it?**
- **Decision – try to test grammar from 2<sup>nd</sup> experiment under the test corpus**

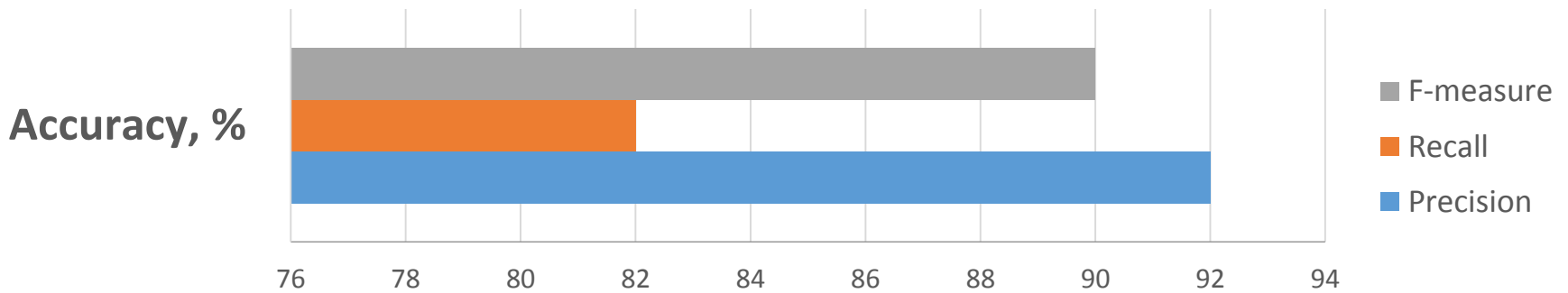
										812	583	168				
										ВЫДАЛЯЕМ ПАУТОРЫ						
										клеім назву для ўнікальнасці	выводзім толькі 1 раз ую назву	маркер	хвосьцік	колькасьць выбранага	сапраўды POS	сапраўды NEG
ДАДЗЕНЬЯ З КАНКАРДАНСА НУДЖ																
neg	1.not	туалетах стоялі не текущие	POS	дырки в полу'. Далеко	neg1.not	neg1.not	neg	1.not		1	0	0				
neg	1.not	обычные унитаз дырки в полу	NEG	. Далеко не все ваши	neg1.not					0	0	0				
neg	10.not	Наверное, путе не стоит	NEG	посещать некоторые и	neg10.not	neg10.not	neg	10.not		1	0	1				
neg	10.not	из помещений м очень крутые	NEG	подъемы с высокими	neg10.not					0	0	0				
neg	10.not	К ним очень кру с высокими ступ	NEG	. GSPLIT:u AlenaKM	neg10.not					0	0	0				
neg	100.not	С одной стороны колхоз	NEG	, когда все по команд	neg100.not	neg100.not	neg	100.not		1	0	1				
neg	100.not	мало. Без маши доехать сложно	NEG	, и второй замок - Нес	neg100.not					0	0	0				
neg	101.not	Но душевности не :	NEG	. Коммерция в чистом	neg101.not	neg101.not	neg	101.not		1	0	1				
neg	101.not	Но душевности Коммерция	NEG	в чистом виде GSP	neg101.not					0	0	0				
neg	101.not	-E9FDC8E34822 )	POS	e	neg101.not					0	0	0				
neg	102.not	Даже слишком хорош	NEG	, как заново построени	neg102.not	neg102.not	neg	103.not		1	0	1				
neg	103.not	экспонатов не м	NEG	, они разрозненны и и	neg103.not	neg103.not	neg	103.not		1	0	1				
neg	103.not	экспонатов не м разрозненны	NEG	и иногда создается вг	neg103.not					0	0	0				
neg	104.not	Что касается пе не все одинаков	NEG	и обходительны, а ук	neg104.not	neg104.not	neg	104.not		1	0	1				
neg	104.not	с собой, на терп цены необьясни	NEG	GSPLIT:u AlenaKMirsk	neg104.not					0	0	0				

# Charts for the 1<sup>st</sup> and the 2<sup>nd</sup> experiments data analysis

## Improving of "right answers" subgraphs



## Accuracy



# Evaluation of 2<sup>nd</sup> grammar under the test corpus

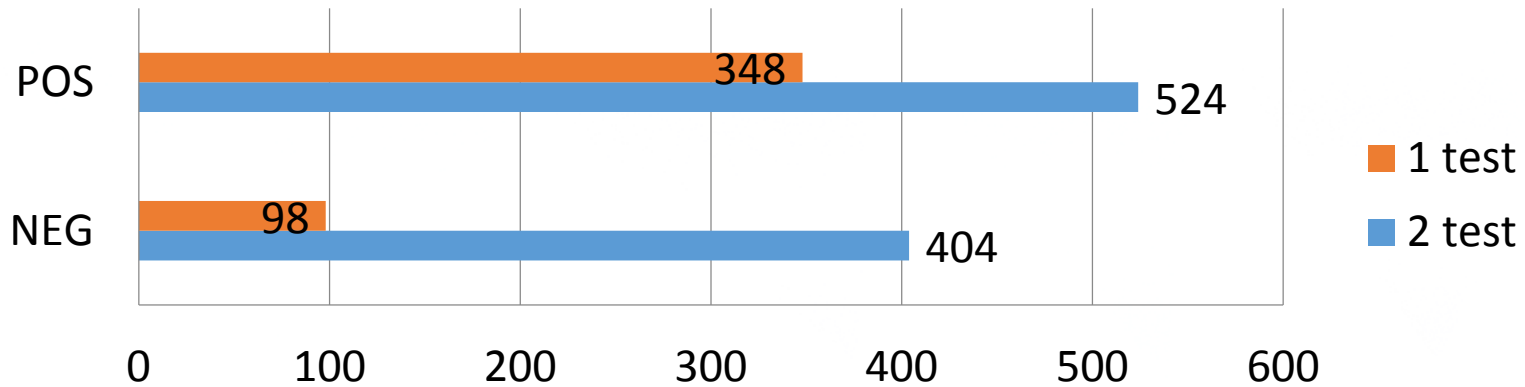
- Precision: 0,94. Recall: 0,70. F-measure: **0,80**
- **Discussion – recall is going to down, but precision – to up. F-measure – is going to down but it is still enough.**
- **Strategic decision – to do**
  - **visualization of all data in a charts**
  - **analytical plans for the stabilization of accuracy**

						ВЫДАЛЯЕМ ПАУТОРЫ				79	68	7
ДАДЗЕНЬЯ З КАНКАРДАНСА НУДЖ						клеім назву для ўнікальнасьці	выводзім толькі 1 раз усю назву	маркер	хвосьцік	колькасьць выбранага	сапраўды POS	сапраўды NEG
neg	1.not	как-то	не очень понрав	NEG	в Мирском	neg1.not	neg1.not	neg	1.not	1	0	0
neg	1.not	кафешек и	недорогой	POS	гостиницы. Печалька	neg1.not				0	0	0
neg	15.not		Очень крутые л	NEG	, запарыцца, пока	neg15.not	neg15.not	neg	18.not	1	0	1
neg	18.not		Минус	NEG	Мирского замка	neg18.not	neg18.not	neg	18.not	1	0	1
neg	18.not	до него	не очень удобн	NEG	добираться из	neg18.not				0	0	0
neg	19.not	чувствуется ста		POS	GSPLIT:u	neg19.not	neg19.not	neg	2.not	1	0	0
...	...	...	...	POS	GSPLIT:u	...	...	...	...	1	0	0

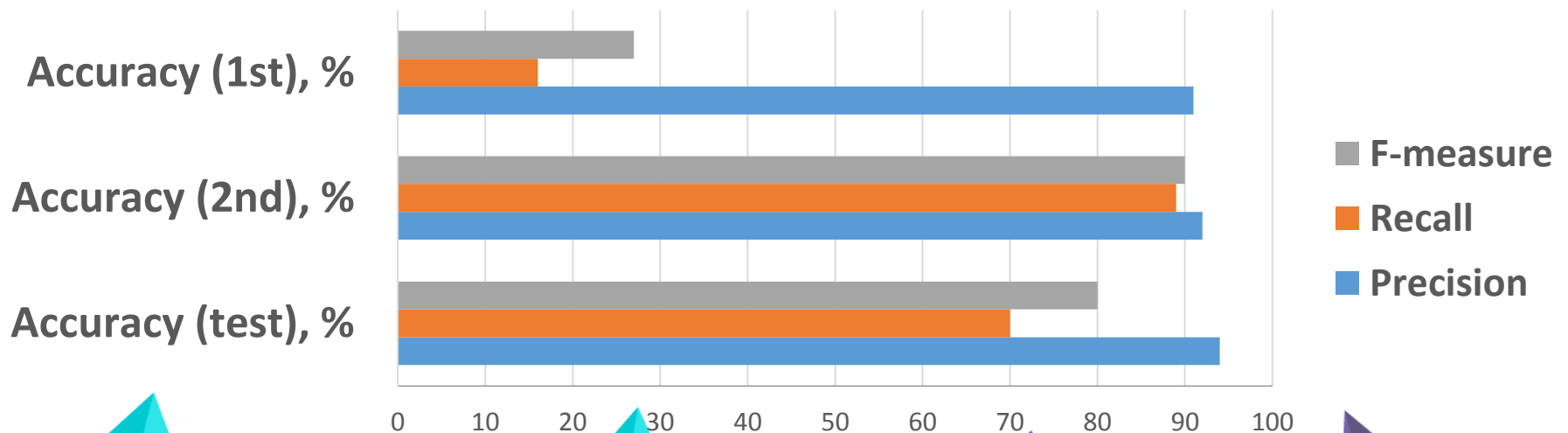


# Charts for the 1<sup>st</sup>, 2<sup>nd</sup> and test experiments data analysis

## Improving of “right answers” in subgraphs



## Improving of accuracies between experiments



# Plans for SA grammars improving & using in applications

- to add <WF> and <P> modifications into grammars for more accurate work grammars in general case
- to improve the constructed syntax grammars algorithm with adding the to a subgraphs of tagged true answers for Adverbs and Adjectives
- to do modifications in Excel calculations when we obtains multiple POS (NEG) annotations for the same review
- to add a tag for neutral state
- to build online Java Belarusian NooJ prototype for sentiment analysis tourism task automatization





# Implementation of new Java NooJ core

# Statement of the problem

- to develop new dictionary module for NOOJ
- to develop new RegExp module for NOOJ
- to test it under the test set that includes short text, 6 tasks for NooJ Linguistic Analysis & Locate pattern functionality of NooJ formalism



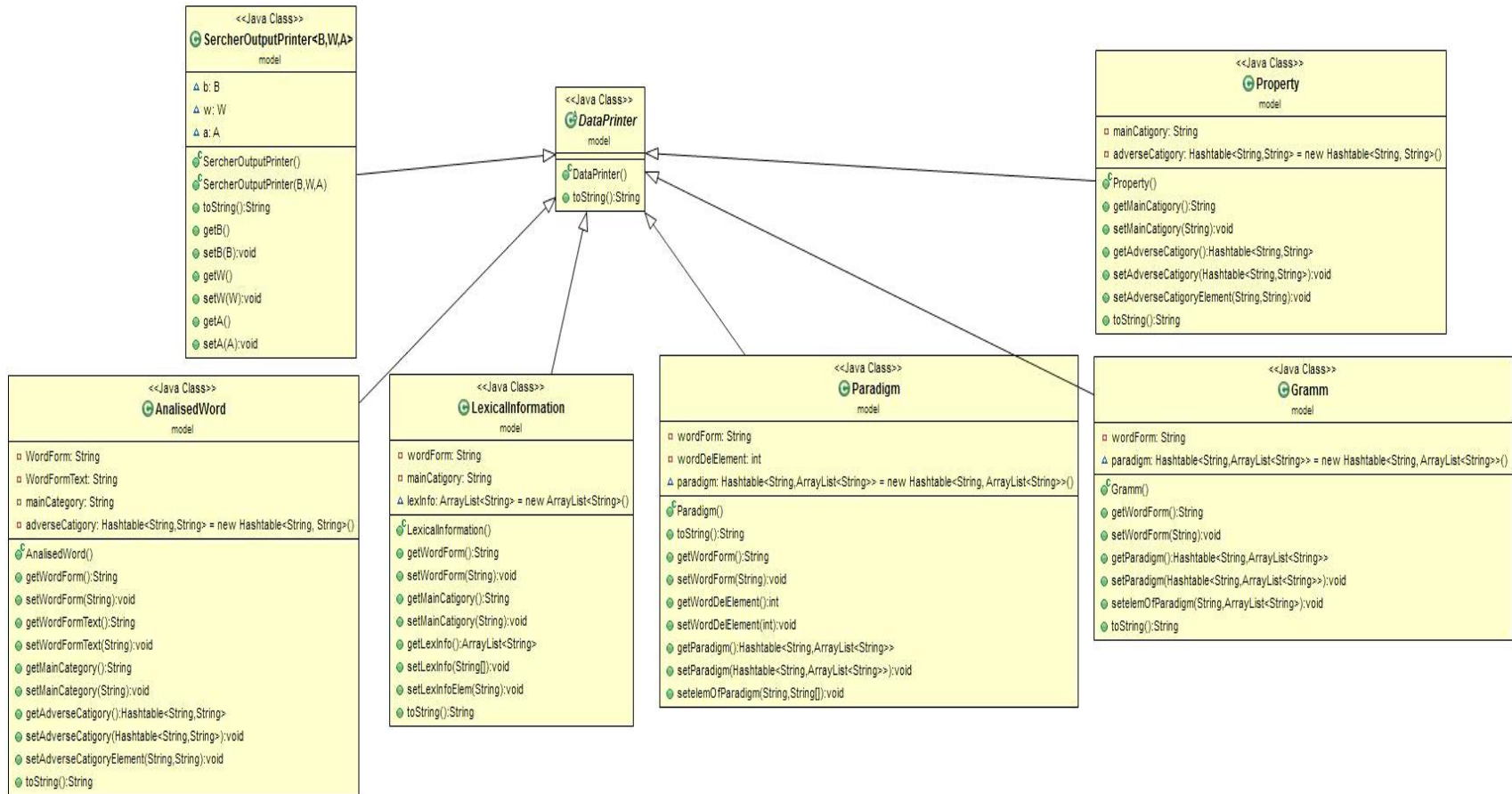


# What we can use?

- C# NOOJ version as working program
- Open source of [Java NOOJ](#) from META-SHARE for modification
- NooJ manual as hand book of NooJ formalizm
- Sample files of \*.dic, \*.nof, \_properties from the Belarusian NooJ module
- Model-View-Controller approche
- E-mails to Max :)



# Data model classes diagram

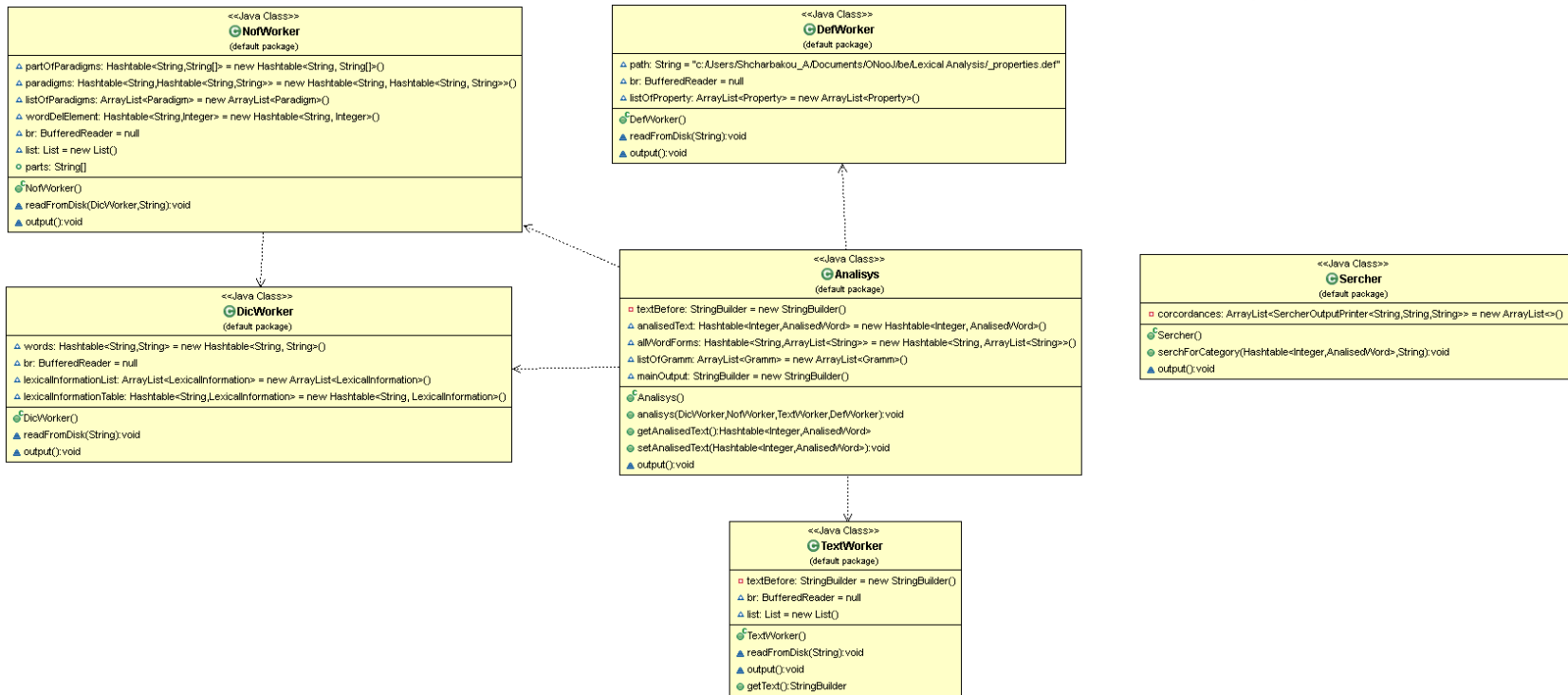


# View

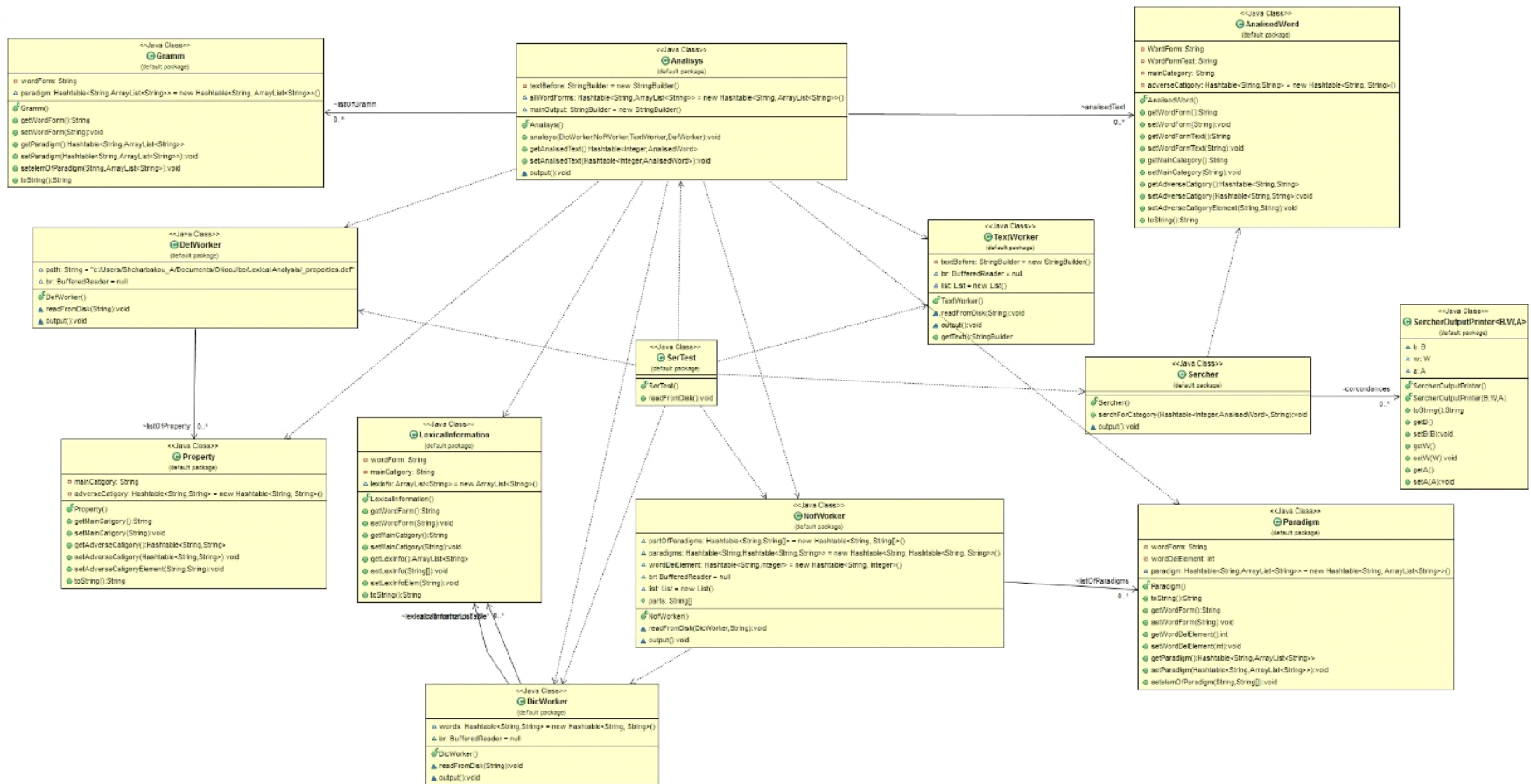
- Simple console output



# Controller classes diagram



# Controller and data models classes diagram



# Descriptions of the developed classes

- **Class AnalisedWord**, are presented in the class diagram implements a full linguistic information stored analyzed word contains fields such as: word form, a particular form of words in the text, some of the language and other linguistic information. Appropriate methods are set and returns the value fields of the class. Just overridden method in the object recording line
- **Class LexicalInformation** shown in the class diagram implements a data model and stores the basic linguistic information word as described in the dictionary. It includes fields such as: word form, part of speech, and other linguistic information. Appropriate methods are set and returns the value fields of the class. Just overridden object recording method in a row.
- **Class Paradigm** shown in the class diagram implements a data model for storage in the computer memory all inflected grammar words. It includes fields such as: word form, the number of changes the sign at the end of words, the variable part of the word. Appropriate methods are set and returns the value fields of the class. Just overridden method in the object recording line



- **Class Property** shown in the class diagram implements a data model and stores the additional linguistic information. It includes fields such as: part of the language and other linguistic information. Appropriate methods are set and returns the value fields of the class. Just overridden object recording method in a row.
- **Class SercherOutputPrinter** shown in the class diagram implements a data model and stores the context in which the word is used in the text. It includes fields such as: context before the word, the word, the context after the word. Appropriate methods are set and returns the value fields of the class. Just overridden method in the object recording line
- **Class Gramm** presented data model to store in the computer memory all inflected grammar words. It includes fields such as: word form, all forms of words with corresponding linguistic information. Appropriate methods are set and returns the value fields of the class. Just overridden method in the object recording lineBelow The the group of classes that belonging to the data controlers : DefWorker, DicWorker, NofWorker, Sercher, Analisis.



- **TextWorker class** provides methods that implement the file processing function of the text type files.
- **Analysis class** provides an implementation of secondary processing of the text prior to analysis, the construction of grammars, text analysis, the preparation of the analysis results to the output file, the output of the analysis of the text in the file.



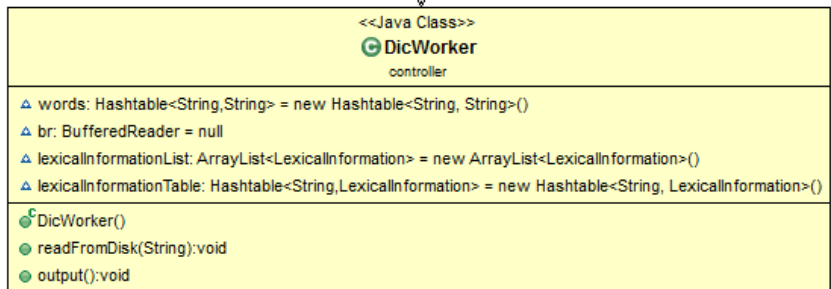
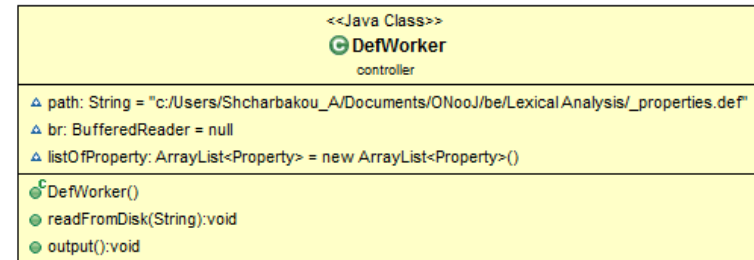
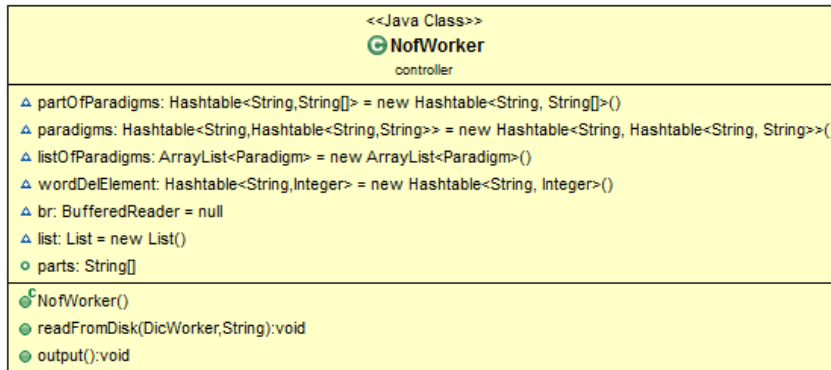


# For the analysis of the text are used 3 types of files for dictionary

- «Dictionary»
- «Inflectional / Derivational description»
- «Dictionary properties definition»



# 3 classes have been implemented for processing dictionaries

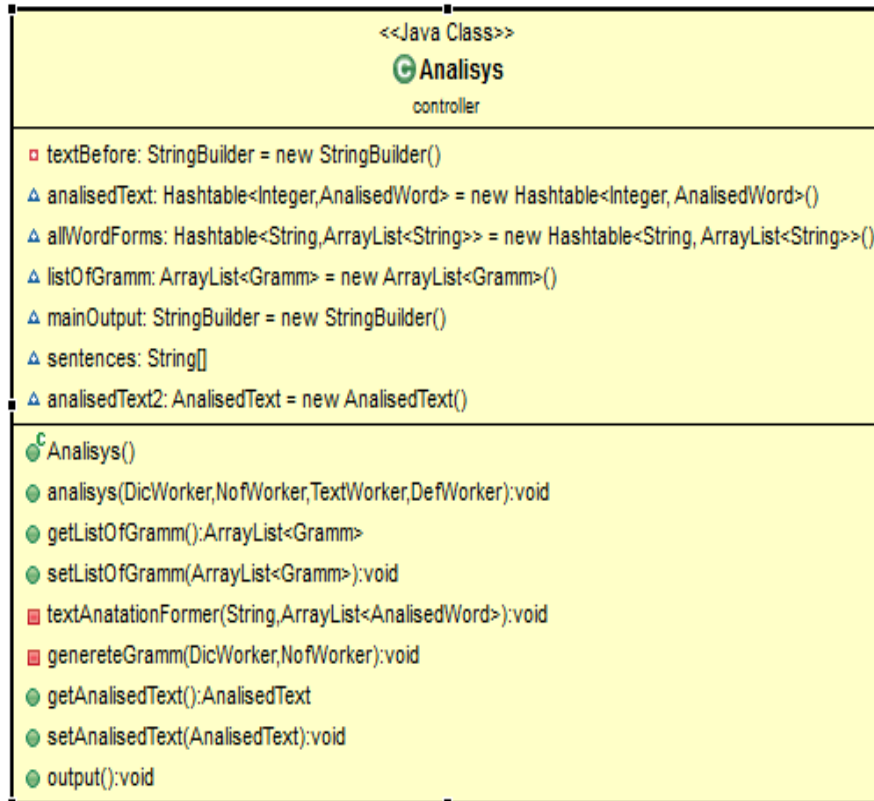


# Description of the developed classes for work with dictionaries

- **DefWorker class** provides methods that implement the file processing function of the type «Dictionary properties definition».
- **DicWorker class** provides methods that implement the file processing function of the type «Dictionary».
- **NofWorker class** provides methods that implement the file processing function of the type «Inflectional / Derivational description».



# Analysis class was implemented for word processing



**Analysis class** provides an implementation of secondary processing of the text prior to analysis, the construction of grammars, text analysis, the preparation of the analysis results to the output file, the output of the analysis of the text in the file.

# One simple example how to work new RegExp

Task: Locate pattern: <VERB+Present>

For text: актрыса сядзела і глядзела на ўваход. актрыс з аб'езду ішла, для таго каб ускочыць. актрысы ідуць на ўваходы тэатраў.

Must be

Text	Before	Seq.	After
	для таго каб ускочыць.	актрысы ідуць	на ўваходы тэатраў.

Answer is:

```
Problems @ Javadoc Declaration Search Console Debug
<terminated> Rerun view.SerTest.readFromDisk [JUnit] C:\Program Files (x86)\Java\jre1.8.0_131\bin\javaw.exe (15 мая 2017 г., 17:10:41)
SercherOutputPrinter [before=для таго каб ускочыць. актрысы , word=ідуць, afte =на ўваходы тэатраў.
```

Approve: yes

# Testing

Task 1. Locate pattern: <NOUN>

Text: актрыса сядзела і глядзела на ўваход.

актрыс з аб'езду ішла, для таго каб ускочыць.

Text	Before	Seq.	After
		Актрыса	сядзела і глядзела на ўваход
Актрыса сядзела і глядзела на	ўваход	.	Актрыс з аб'езду ішла
сядзела і глядзела на ўваход.	Актрыс	з аб'езду ішла, для	
глядзела на ўваход. Актрыс з	аб'езду	ішла, для таго каб ускочыць	

Answer is:

```
Problems @ Javadoc Declaration Search Console Debug
<terminated> Rerun view.SerTest [JUnit] C:\Program Files (x86)\Java\jre1.8.0_131\bin\javaw.exe (15 мая 2017 г., 16:54:34)
SercherOutputPrinter [before=, word=актрыса, after=сядзела і глядзела на ]
SercherOutputPrinter [before=актрыса сядзела і глядзела на , word=ўваход., after=актрыс з аб'езду ішла, ]
SercherOutputPrinter [before=сядзела і глядзела на ўваход. , word=актрыс, after=з аб'езду ішла, для ]
SercherOutputPrinter [before=глядзела на ўваход. актрыс з , word=аб'езду, after=ішла, для таго каб ]
```

Approve: yes. Comments: there are some mistakes in letters case



# Task 2. Locate pattern: < VERB >

**Text:** актрыса сядзела і глядзела на ўваход.  
актрыс з аб'езду ішла, для таго каб ускочыць.

**Answer must be:**

Text	Before	Seq.	After
	Актрыса	сядзела	і глядзела на ўваход. Актрыс
	Актрыса сядзела і	глядзела	на ўваход. Актрыс з аб
	ўваход. Актрыс з аб'езду	ішла	, для таго каб ускочыць.
	'езду ішла, для таго каб	ускочыць	.

**Answer is:**

```
Problems @ Javadoc Declaration Search Console Debug
<terminated> Rerun view.SerTest.readFromDisk [JUnit] C:\Program Files (x86)\Java\jre1.8.0_131\bin\javaw.exe (15 мая 2017 г., 17:02
SercherOutputPrinter [before=актрыса , word=сядзела, after=і глядзела на ўваход. ]
SercherOutputPrinter [before=актрыса сядзела і , word=глядзела, after=на ўваход. актрыс з ]
SercherOutputPrinter [before=на ўваход. актрыс з аб'езду , word=ішла,, after=для таго каб ускочыць
SercherOutputPrinter [before=аб'езду ішла, для таго каб , word=ускочыць., after=]
```

**Approve:** yes. **Comments:** please check small mistakes in punctuation

# Task 3. Locate pattern: <актрысамі>

- Text: актрыса сядзела і глядзела на ўваход. актрыс з аб'езду ішла, для таго каб ускочыць.
- Answer must be:

Text	Before	Seq.	After
		Актрыса	сядзела і глядзела на ўваход
сядзела і глядзела на ўваход.	Актрыс		з аб'езду ішла, для

- Answer is:

```
Problems @ Javadoc Declaration Search Console Debug
<terminated> Rerun view.SerTest.readFromDisk [JUnit] C:\Program Files (x86)\Java\jre1.8.0_131\bin\javaw.exe (15 мая 2017 г., 17:06:04)
SercherOutputPrinter [before=сядзела і глядзела на ўваход. , word=актрыс, after=з аб'езду ішла, для ]
SercherOutputPrinter [before=, word=актрыса, after=сядзела і глядзела на ]
```

- Approve: yes. Comments: please check case of letters





## Task 4. Locate pattern: <актрысамі><VERB>

- Text: актрыса сядзела і глядзела на ўваход. актрыс з аб'езду ішла, для таго каб ускочыць.
- Answer must be:

Text	Before	Seq.	After
	Актрыса сядзела	і глядзела на ўваход.	Актрыс

- Answer is:

```
Problems @ Javadoc Declaration Search Console Debug
<terminated> Rerun view.SerTest.readFromDisk [JUnit] C:\Program Files (x86)\Java\jre1.8.0_131\bin\javaw.exe (15 мая 2017 г., 17:10:41)
SercherOutputPrinter [before=для таго каб ускочыць. актрысы , word=ідуць, after=на ўваходы тэатраў. ]
```

- Approve: no. Comments: must be other word located.



## Task 5. Locate pattern: <VERB+Present>

For text: актрыса сядзела і глядзела на ўваход. актрыс з аб'езду ішла, для таго каб ускочыць. актрысы ідуць на ўваходы тэатраў.

Must be

Text	Before	Seq.	After
для таго каб ускочыць. актрысы ідуць на ўваходы тэатраў.			

Answer is:

```
Problems @ Javadoc Declaration Search Console Debug
<terminated> Rerun view.SerTest.readFromDisk [JUnit] C:\Program Files (x86)\Java\jre1.8.0_131\bin\javaw.exe (15 мая 2017 г., 17:10:41)
SercherOutputPrinter [before=для таго каб ускочыць. актрысы , word=ідуць, afte =на ўваходы тэатраў.
```

Approve: yes

Task 6. Locate pattern: <актрысы><VERB+Present>

- For text from task 5
- Answer must be:

Text	Before	Seq.	After
	ішла, для таго каб ускочыць.	актрысы ідуць	на ўваходы тэатраў.

- Answer is:

```
Problems @ Javadoc Declaration Search Console Debug
<terminated> Rerun view.SerTest.readFromDisk [JUnit] C:\Program Files (x86)\Java\jre1.8.0_131\bin\javaw.exe (15 мая 2017 г., 17:13:19)
SercherOutputPrinter [before=ішла, для таго каб ускочыць. , word=актрысы ідуць, after=ідуць на ўваходы тэатраў. ]
```

- Approve: yes.

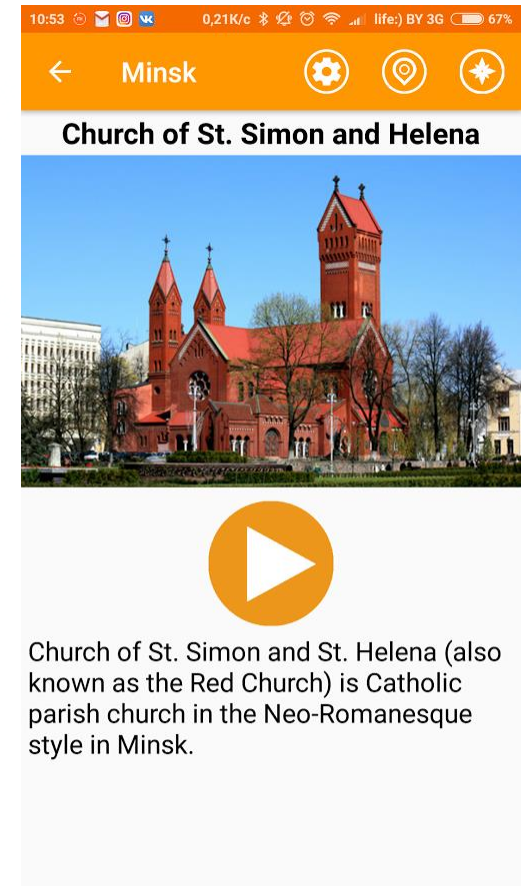
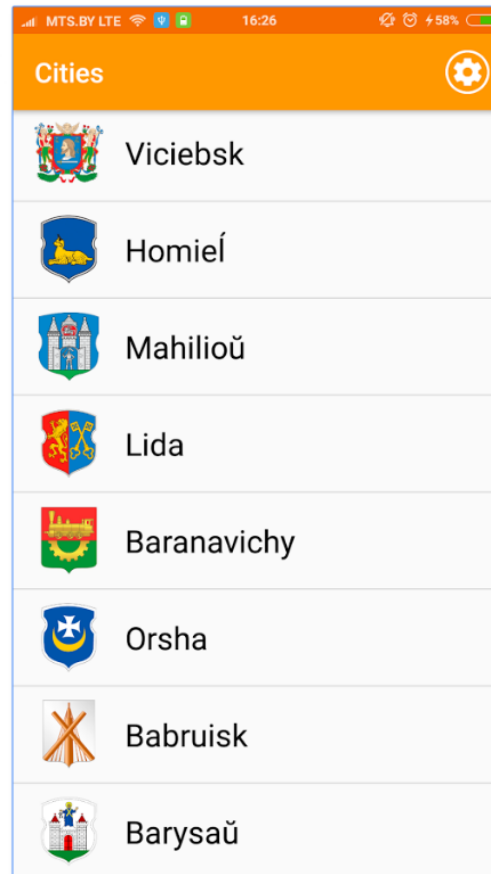
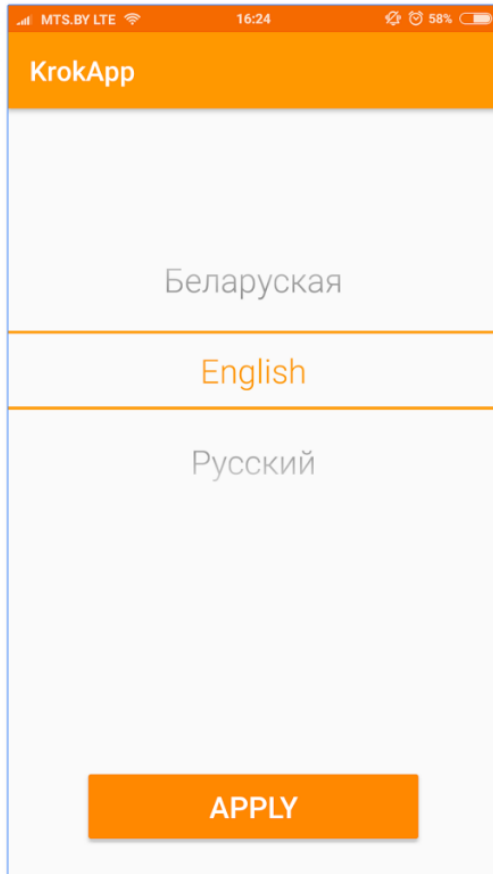
# Plans for further development of new core of Java NooJ

- To complete the RegExp module. Re-check tests.
- To finalize the dictionary module. Make tests.
- To extend the functionality for NooJ grammars.





# KrokApp application in touristic sphere



## **Yuras Hetsevich**

PhD, Head of the Laboratory of Speech  
Synthesis and Speech Recognition UIIP  
NAS Belarus

yuras.hetsevich@gmail.com  
+375 29 681 52 53  
Belarus, Minsk



**Thank you for your attention!**

شكرا جزيلاً

Questions? Suggestions? Cooperation?

